Topological insulators beyond the Brillouin zone via Chern parity

Andrew M. Essin[1] and J. E. Moore[1, 2]

[1]*Department of Physics, University of California, Berkeley, CA 94720*
[2]*Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720*

(Dated: February 5, 2008)

The topological insulator is an electronic phase stabilized by spin-orbit coupling that supports propagating edge states and is not adiabatically connected to the ordinary insulator. In several ways it is a spin-orbit-induced analogue in time-reversal-invariant systems of the integer quantum Hall effect (IQHE). This paper studies the topological insulator phase in disordered two-dimensional systems, using a model graphene Hamiltonian introduced by Kane and Mele as an example. The nonperturbative definition of a topological insulator given here is distinct from previous efforts in that it involves boundary phase twists that couple only to charge, does not refer to edge states, and can be measured by pumping cycles of ordinary charge. In this definition, the phase of a Slater determinant of electronic states is determined by a Chern parity analogous to Chern number in the IQHE case. Numerically we find, in agreement with recent network model studies, that the direct transition between ordinary and topological insulators that occurs in band structures is a consequence of the perfect crystalline lattice. Generically these two phases are separated by a metallic phase, which is allowed in two dimensions when spin-orbit coupling is present. The same approach can be used to study three-dimensional topological insulators.

PACS numbers: 73.43.-f, 85.75.-d, 73.20.At, 73.20.Fz, 72.25.-b, 03.65.Vf

I. INTRODUCTION

Considerable theoretical and experimental effort has been devoted to the quest for an intrinsic spin Hall effect[1,2,3,4] that would allow generation of spin currents by an applied electric field. Interesting mechanisms for such spin current generation make use of spin-orbit coupling, which breaks the $SU(2)$ spin symmetry of free electrons but not time-reversal symmetry. A dissipationless type of intrinsic spin Hall effect was predicted[5,6] to arise in materials that have an electronic energy gap. This "quantum spin Hall effect" (QSHE) in certain materials with time-reversal symmetry has a subtle relationship to the integer quantum Hall effect, in which time-reversal symmetry is explicitly broken by a magnetic field.

In a system with unbroken time-reversal symmetry, a dissipationless charge current is forbidden, but a dissipationless transverse spin current is allowed, of the form

$$\mathcal{J}^i_j = \alpha \epsilon_{ijk} E_k. \tag{1}$$

The current on the left is a spin current and ϵ is the fully antisymmetric tensor. Note that a spin current requires two indices, one for the direction of the current and one for the direction of angular momentum that is transported. The constant of proportionality α depends on the specific mechanism: for example, the (dissipative) extrinsic D'yakonov-Perel mechanism[7] predicts a small α that depends on impurity concentration. The QSHE builds on the construction by Haldane[8] of a lattice "Chern insulator" model, with broken time-reversal symmetry but without net magnetic flux, that shows a $\nu = 1$ IQHE. The simplest example of a QSHE is obtained by taking two copies of Haldane's model, one for spin-up electrons along some axis and one for spin-down. Time-reversal symmetry can be maintained if the effective IQHE magnetic fields are opposite for the two spin components. Then an applied electric field generates a transverse current in one direction for spin-up electrons, and in the opposite direction for spin-down electrons. There is no net charge current, consistent with time-reversal symmetry, but there is a net spin current. However, models like this in which one component of spin is perfectly conserved are both unphysical, since realistic spin-orbit coupling does not conserve any component, and not very novel, since for each spin component the physics is exactly the same as Haldane's model and the spin components do not mix.

More subtle physics emerges when one asks how the QSHE appears in more realistic band structures. Remarkably, band insulators of noninteracting two-dimensional electrons with spin-orbit coupling divide into two classes: the "ordinary insulator", which in general has no propagating edge modes and no spin Hall effect, and the "topological insulator", which has stable propagating edge modes and a generic spin Hall effect, although the amount of spin transported (the coefficient α in (1)) is nonuniversal. These phases are associated with a $\mathbb{Z}_2$-valued topological invariant (an "oddness" or "evenness", in the language of parity*) in the same way that IQHE phases are associated with an integer-valued topological invariant. For explicitness, consider the model of graphene introduced by Kane and Mele.[5] This is a tight-binding model for independent electrons on the honeycomb lattice (Fig. 1). The spin-independent part of the Hamiltonian consists of a nearest-neighbor hopping, which alone would give a semimetallic spectrum

* Throughout "parity" is used to denote oddness or evenness, rather than a spatial inversion eigenvalue.

with Dirac nodes at certain points in the 2D Brillouin zone, plus a staggered sublattice potential whose effect is to introduce a gap:

$$H_0 = t \sum_{\langle ij\rangle\sigma} c^\dagger_{i\sigma} c_{j\sigma} + \lambda_v \sum_{i\sigma} \xi_i c^\dagger_{i\sigma} c_{i\sigma}. \tag{2}$$

Here $\langle ij\rangle$ denotes nearest-neighbor pairs of sites, σ is a spin index, ξ_i alternates sign between sublattices of the honeycomb, and t and λ_v are parameters.

The insulator created by increasing λ_v is an unremarkable band insulator. However, the symmetries of graphene also permit an "intrinsic" spin-orbit coupling of the form

$$H_{SO} = i\lambda_{SO} \sum_{\langle\langle ij\rangle\rangle\sigma_1\sigma_2} \nu_{ij} c^\dagger_{i\sigma_1} s^z_{\sigma_1\sigma_2} c_{j\sigma_2}. \tag{3}$$

Here $\nu_{ij} = (2/\sqrt{3})\hat{\boldsymbol{d}}_1 \times \hat{\boldsymbol{d}}_2 = \pm 1$, where i and j are next-nearest-neighbors and $\hat{\boldsymbol{d}}_1$ and $\hat{\boldsymbol{d}}_2$ are unit vectors along the two bonds that connect i to j. The Hamiltonian $H_0 + H_{SO}$ conserves s^z, the distinguished component of electron spin, and reduces for fixed spin (up or down) to Haldane's model.[8]

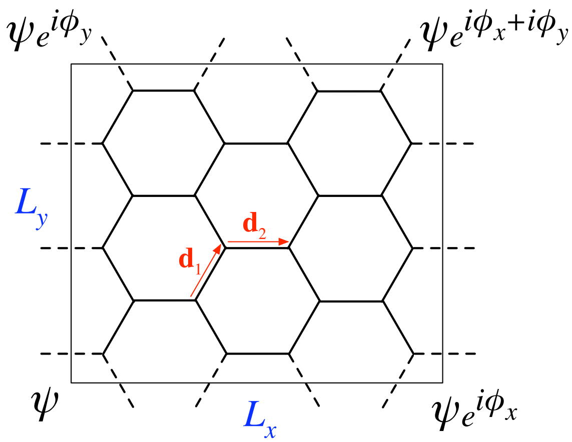

FIG. 1: The honeycomb lattice on which the tight-binding Hamiltonian resides. For the two sites depicted, the factor ν_{ij} of equation (3) is $\nu_{ij} = -1$. The phases $\phi_{x,y}$ describe twisted boundary conditions, introduced in equation (10).

The "topological insulator" phase created when $|\lambda_{SO}| \gg |\lambda_v|$ is quite different from the ordinary insulator that appears when $|\lambda_v| \gg |\lambda_{SO}|$ (here we assume that there is an energy gap between the lower and upper band pairs in which the Fermi level lies). The former has counterpropagating edge modes and shows the QSHE, while the latter does not.[5] Does this phase exist for more realistic spin-orbit coupling? The spin component s_z is no longer a good quantum number when the Rashba spin-orbit coupling is added:

$$H_R = i\lambda_R \sum_{\langle ij\rangle\sigma_1\sigma_2} c^\dagger_{i\sigma_1} \left(\boldsymbol{s}_{\sigma_1\sigma_2} \times \hat{\boldsymbol{d}}_{ij}\right)_z c_{j\sigma_2}, \tag{4}$$

with $\boldsymbol{d}_{ij}$ the vector from $i \to j$ and $\hat{\boldsymbol{d}}_{ij}$ the corresponding unit vector. (Note that Rashba spin-orbit coupling is not intrinsic to graphene but generated by inversion-symmetry breaking in the out-of-plane direction.) The topological insulator survives but is strongly modified in the presence of this term. For a general 2D band structure with s^z conserved, there are many phases labeled by an integer n, as in the IQHE: if spin-up electrons are in the $\nu = n$ state, then spin-down electrons must be in the $\nu = -n$ state by time-reversal symmetry, where the sign indicates that the direction of the effective magnetic field is reversed. Once s^z is not conserved, there are only two insulating phases, the "ordinary" and "topological" insulators. A heuristic definition of the topological insulator, without reference to any particular spin component or the spin Hall effect, is as a band insulator that is required to have gapless propagating edge modes at the sample boundaries. The decoupled $\nu = \pm n$ cases with s^z conserved are adiabatically connected, once s^z is not conserved, to the ordinary insulator for even n and to the topological insulator for odd n. A review of how these two cases emerge as the only possibilities in 2D follows in Section II.

It is not obvious at first glance how to generalize the topological insulator phase to finite, noncrystalline systems, rather than band structures, as when the parameters of the Hamiltonian $H = H_0 + H_{SO} + H_R$ are drawn from a random distribution. The first approach was in terms of a spin Chern number[9] similar to the Chern integer in finite IQHE systems, but there is now agreement that for a clean band structure the only invariants are of $\mathbb{Z}_2$ type, rather than integer type.[10,11,12] Two equivalent definitions of the appropriate $\mathbb{Z}_2$ invariant for a finite disordered system, in the simple case when the disorder splits all degeneracies other than Kramers degeneracies, are as follows (the full definition is given and compared to previous work in the following section). The finite system can be considered as a unit "supercell" of a large 2D lattice. A large, finite supercell gives many bands, but each pair of bands connected by time reversal (Kramers pair) can be assigned its own $\mathbb{Z}_2$ invariant.[11] The phase of the supercell system, if the Fermi level lies in a gap, is then identified by adding up all the invariants (mod 2). Alternately, a direct definition of the phase in the finite system can be given that is related to the notion of "$\mathbb{Z}_2$ pumping".[10] Real charge is pumped as the flux through the system is taken from 0 to $hc/2e$ (*half* the usual flux quantum that appears in IQHE pumping); we show that in the topological insulator, any pumping cycle, properly defined, pumps an odd number of electron charges, while for the ordinary insulator any cycle pumps an even number of charges.

We implement this definition numerically using an explicit algorithm introduced by Fukui and Hatsugai[13] for computing $\mathbb{Z}_2$ topological invariants on a Brillouin zone. The topological insulator phase is robust to disorder: while different realizations of disorder assign different "Chern parities" to individual subbands, it is found that

the total for occupied subbands is always "odd" for a wide range of parameters, which in our definition indicates a topological insulator. In the IQHE, a pair of bands of opposite Chern number can annihilate as the strength of disorder is increased; in the QSHE, two band pairs that both have odd Chern parity can annihilate, i.e., become two even-parity band pairs. If the topological insulator can be destroyed by band annihilation, then there are extended (i.e., topologically nontrivial) states with an arbitrarily small gap; it may be the case that for some range of parameters, there are extended states at the Fermi level even in the thermodynamic limit, indicating a metallic phase. In the IQHE, there is only a single energy with extended states rather than a range of energies, and hence no metallic phase. We find the phase diagram of the graphene model with on-site disorder, and in the presence of non-zero Rashba coupling find evidence for a metallic phase intervening between ordinary and topological insulators.

Recent work by Obuse *et al.*[14] obtains a phase diagram and critical exponents using a network model for the spin quantum Hall effect that is similar to the Chalker-Coddington network model[15] for the IQHE (see also Onoda *et al.*[16] for a quasi-1D study of localization in the Kane-Mele Hamiltonian with disorder). Our results on the phase diagram are consistent with these works, although our method is unable to generate large enough system sizes to confirm the exponents found for the phase transitions. To understand how the network and Chern-parity approaches complement each other, consider the integer quantum Hall effect (IQHE): while the phenomenological network approach to the IQHE is valuable both to find the critical indices precisely and to identify the minimal necessary elements of a theory for the transition, Chern-number studies remain important for studies of effects such as the floating of extended states,[17,18,19] where knowledge of the topological properties of a state is required. The network model gives more accurate information about the phase transitions but, if only the localization length is probed, does not distinguish the different phases in bulk. A more technical difference between the two approaches is discussed at the end of Section II.

Section II reviews how the topological insulator phase in perfect crystals arises from a parity-valued topological invariant of the band structure, similar to the TKNN[20] integers in the IQHE. It then gives two mathematically equivalent definitions of the topological insulator phase in disordered systems based on Chern parity. One definition simply considers a finite disordered system as a supercell of an infinite lattice system, while the other is based upon closed charge pumping cycles driven by application of flux to a finite periodic system. Section III reviews the Fukui-Hatsugai algorithm[13] adapted to numerical computation of these invariants in disordered systems, then computes the phase diagram of the Kane-Mele graphene model[5] with on-site disorder. The conclusions of our study for general 2D disordered systems are summarized in a short Section IV. While there is a three-dimensional version of the QSHE[11,21,22,23] that is less directly connected to the IQHE and has interesting localization behavior, we will restrict our attention to 2D except for some comments in the final section.

II. CHERN PARITIES FOR DISORDERED NONINTERACTING ELECTRON SYSTEMS

A. The definition of the $\mathbb{Z}_2$ invariant in clean systems

We review one definition of the $\mathbb{Z}_2$ invariant of a band pair in a 2D band structure, then explain its generalization to noncrystalline systems. With the Hamiltonian

$$H = H_0 + H_{SO} + H_R \tag{5}$$

defined in equations (2), (3), and (4), or any periodic, single-electron Hamiltonian, a Berry connection $\boldsymbol{\mathcal{A}}$ can be defined on the Brillouin zone (BZ) from the periodic part $u(\boldsymbol{k})$ of a Bloch state $\psi_{\mathbf{k}} = u(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}$. For a single nondegenerate band, we define $\boldsymbol{\mathcal{A}}_j(\boldsymbol{k}) = -i\langle u_j|\boldsymbol{\nabla}_{\boldsymbol{k}}|u_j\rangle$ for band j. This Berry connection can be used to understand how a commensurate magnetic field in a two-dimensional lattice system, which allows definition of Bloch states on the (magnetic) Brillouin zone, leads to the IQHE. The TKNN integers[20] in the IQHE can be written as integrals over the BZ of the Berry field strength $\mathcal{F} = (\boldsymbol{\nabla}_{\boldsymbol{k}} \times \boldsymbol{\mathcal{A}})_z$,

$$\begin{aligned} n_j &= -\frac{1}{2\pi}\int_{BZ} \mathcal{F}_j\, d^2k \\ &= \frac{i}{2\pi}\int_{BZ}\left[\left\langle \frac{\partial u}{\partial k_x}\middle|\frac{\partial u}{\partial k_y}\right\rangle - \left\langle \frac{\partial u}{\partial k_y}\middle|\frac{\partial u}{\partial k_x}\right\rangle\right]. \end{aligned} \tag{6}$$

The TKNN integers are integer-valued topological invariants: the TKNN integer of a band cannot change as long as the band remains nondegenerate. When bands touch, only the total TKNN integer of the bands is well-defined,[24] and the IQHE state of a gapped system is given by the sum of TKNN integers for occupied bands. However, all these integers vanish in a time-reversal-invariant band structure (see, e.g., Ref. 11). Instead there is a $\mathbb{Z}_2$ invariant associated with a band *pair* in a time-reversal-invariant 2D Fermi system.[6] Time-reversal invariance requires that

$$\Theta\mathcal{H}(-\boldsymbol{k})\Theta^{-1} = \mathcal{H}(\boldsymbol{k}), \tag{7}$$

where $\mathcal{H}(\boldsymbol{k})$ is the Bloch Hamiltonian and $\Theta = -i\sigma_y K$ is the action of time reversal (K performs complex conjugation and σ_y, the usual Pauli matrix, acts on spin indices).

With time-reversal breaking, as in the IQHE, it is reasonable to ignore band degeneracies, but time-reversal invariance forces the single electron energies to be doubly degenerate at certain time-reversal-invariant momenta (see, for example, Ref. 25). These degeneracies are known

as "Kramers degeneracies". The single-band connection $\boldsymbol{\mathcal{A}}_j$ will not be globally defined. So long as each Kramers pair remains separated by gaps in the energy spectrum from all others, the appropriate generalization is

$$\begin{aligned}\boldsymbol{\mathcal{A}}_j &= -i\left(\langle u_{j1}|\boldsymbol{\nabla_k}|u_{j1}\rangle + \langle u_{j2}|\boldsymbol{\nabla_k}|u_{j2}\rangle\right)\\ &= -i\,\mathrm{tr}\,\boldsymbol{u}_j^\dagger \boldsymbol{\nabla_k}\boldsymbol{u}_j. \end{aligned} \tag{8}$$

In the second equation, $\boldsymbol{u}_j = (u_{j1}, u_{j2})$ where the Bloch functions are viewed as column vectors, i.e., $\boldsymbol{u}_j$ is a matrix. This compact notation follows Fukui and Hatsugai[12] and makes it simple to include more than one band pair simply by adding more columns to $\boldsymbol{u}$. Finally, we again define a field strength associated with the potential, $\mathcal{F} = (\boldsymbol{\nabla_k} \times \boldsymbol{\mathcal{A}})_z$.

From these quantities, Fu and Kane[10] give the following formula for the $\mathbb{Z}_2$ topological invariant in terms of the Bloch functions of the clean system:

$$D = \frac{1}{2\pi}\left[\oint_{\partial(EBZ)} d\boldsymbol{k}\cdot\boldsymbol{\mathcal{A}} - \int_{EBZ} d^2k\,\mathcal{F}\right] \mod 2. \tag{9}$$

The notation EBZ stands for Effective Brillouin Zone,[11] which describes one half of the Brillouin zone together with appropriate boundary conditions. Since the BZ is a torus, the EBZ can be viewed as a cylinder, and its boundary ∂(EBZ) as two circles, as in Fig. 2(b). While $\mathcal{F}$ is gauge-invariant, $\boldsymbol{\mathcal{A}}$ is not, and different (time-reversal-invariant) gauges can change the sum of boundary integral by an even amount. Heuristically, (9) states that integrating the field strength of a Kramers pair of bands over half the Brillouin zone gives a topological invariant in the same way as integrating the field strength of a single band over the whole zone. However, a gauge-dependent boundary term must be added to make the integral an integer. As mentioned, this boundary term is ambiguous (via gauge changes) up to an even integer, so only the parity D mod 2 is well-defined.

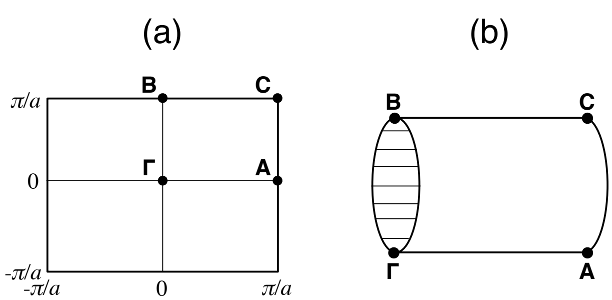

FIG. 2: (a) A two-dimensional Brillouin zone; note that any such Brillouin zone, including that for graphene, can be smoothly deformed to a torus. The labeled points are time-reversal-invariant momenta. (b) The effective Brillouin zone (EBZ). The horizontal lines on the boundary circles ∂(EBZ) connect time-reversal-conjugate points, where the Hamiltonians are related by time reversal and so cannot be specified independently.

A direct proof of the existence and $\mathbb{Z}_2$ nature of the topological invariant for multiple bands without introducing gauge-dependent quantities can be obtained[11] by considering maps, from the EBZ to the space of Bloch Hamiltonians, that are consistent with time-reversal, following work on the IQHE by Avron *et al.*[24] The Berry field strength can be written in terms of the (gauge-invariant) projection operator onto the band pair rather than the (gauge-dependent) wavefunctions. In this approach, the ambiguity by an even integer that is crucial to obtain a $\mathbb{Z}_2$ rather than $\mathbb{Z}$ invariant corresponds to the many different ways in which the circles that form EBZ boundaries in Fig. 2(b) can be contracted to make the EBZ into a sphere. The boundary integrals in (9) just calculate the contribution to the Chern number from these "contractions." On this sphere, Chern integers are well-defined for each nondegenerate band pair, but the different ways of contracting the boundaries cause the resulting integers to differ by even numbers. An explicit numerical implementation of the Fu-Kane formula (9) was given by Fukui and Hatsugai[13] and will be reviewed in Section III.

The invariant D classifies the topology of band pairs; each pair carries a value of $D = 0$ or 1. The model of graphene we study has four bands (two sites per unit cell, two spin states per site), so two band pairs D_1 and D_2, and the total band structure must have $D = 0$ mod 2,[11,24] so that $D_1 = D_2$. Hence there are two phases, characterized by the value of $D = D_1$, say. Now, if at some value of parameters the upper and lower band pairs meet somewhere in the BZ, D is not well-defined, so there can be a phase transition. The two pairs with $D = 1$ annihilate into two $D = 0$ pairs when the bands collide as the parameters are varied. Note that, just as total Chern number is conserved when bands collide in the IQHE, total $\mathbb{Z}_2$ is conserved here.

At half filling, these phases are both insulating, since there is a gap between the second and third bands; the semimetal that appears when the gap closes marks the phase boundary. What happens when disorder is added? We now explain how the definition of the topological insulator generalizes to disordered systems. Just as TKNN integers of a band structure give rise to Chern integers in a disordered system of finite size, the $\mathbb{Z}_2$ invariants of band pairs become "Chern parities".

B. The topological insulator phase for Slater determinants via Chern parity

The TKNN integers generalize in the presence of disorder and interactions to the Chern number of the many-particle wavefunction.[26] A natural question is how disorder and interactions modify the $\mathbb{Z}_2$ invariants of band pairs in spin-orbit-coupled 2D band structures. All derivations of the $\mathbb{Z}_2$ invariants of clean systems depend on Fermi statistics in some way: for example, the existence of Kramers degeneracies and the related fact that

the time-reversal operator squares to -1 both depend on Fermi statistics. A many-fermion wavefunction describing an even number of fermions does not behave in the same way as single-fermion wavefunctions under time-reversal. Hence, given only the many-fermion wavefunction, it does not seem likely that there is a generalization of the $\mathbb{Z}_2$ invariant.

However, for the particular case of many-fermion wavefunctions that are single Slater determinants of single-particle wavefunctions, the invariant can be generalized as we now show. While the assumption of a single Slater determinant limits the treatment of interactions to the Hartree-Fock level, there is no requirement that the wavefunctions in the Slater determinant be Bloch states. As a result, the topological insulator and QSHE can be defined and studied for any disorder strength.

Niu *et al.*[26] and Avron and Seiler[27] showed that for disordered quantum Hall systems there exists a generalization of the TKNN invariant defined for clean systems.[20] They introduce generalized periodic boundary conditions and find an invariant Chern number, similar in form to the TKNN invariant, on the space of boundary phases. Consider a finite system of noninteracting electrons with boundary conditions that are periodic up to phases ϕ_x, ϕ_y, as shown in Fig. 1: this is equivalent to putting magnetic fluxes $\Phi_{x,y} = \phi_{x,y}\Phi_0/2\pi$ through the two noncontractible circles on the torus ($\Phi_0 = hc/e$ is the magnetic flux quantum). As motivation, think of the finite system as a (possibly very large) unit cell of a lattice system. Then in order to determine the phase of this lattice system, instead of integrating over k to do the integrals in the Fu-Kane formula (9), our strategy will be to integrate over the boundary phases, which introduce offsets to the wave vectors. We first carry out this procedure, show that it reproduces the band-structure result for clean systems, and then discuss a physical picture and its relation to previous definitions.

Consider the single-particle wavefunctions of a lattice Hamiltonian such as the graphene model on a finite lattice of size $L_x \times L_y$ (see Fig. 1). Instead of the physical boundary conditions $\psi(\boldsymbol{x} + \boldsymbol{L}_{x,y}) = \psi(\boldsymbol{x})$ for a single-particle wavefunction ψ, introduce the boundary phases, or "twists", $\boldsymbol{\phi} = (\phi_x, \phi_y)$ via

$$\psi(\boldsymbol{x} + \boldsymbol{L}_x) = e^{i\phi_x}\psi(\boldsymbol{x}),\ \psi(\boldsymbol{x} + \boldsymbol{L}_y) = e^{i\phi_y}\psi(\boldsymbol{x}). \quad (10)$$

In the special case of a clean system, this shifts each allowed value of wave vector $\boldsymbol{k}$ by $\boldsymbol{\Delta} = (\phi_x/L_x, \phi_y/L_y)$. Because it is simpler numerically to work with single-valued wavefunctions than the multiple-valued ones of equation (10), and because the standard formalism of the Berry connection assumes the parameters to be explicit in the form of the Hamiltonian rather than living in the boundary conditions (i.e., the definition of the Hilbert space), we want to transfer the twist angles to the Hamiltonian. This is done by making a unitary transformation to

$$\chi(\boldsymbol{x}) = e^{-i(\phi_x x/L_x + \phi_y y/L_y)}\psi(\boldsymbol{x}). \quad (11)$$

Note that for the special case $\phi_{x,y} = 2\pi$ the complex phase in (11) is single-valued as we translate the coordinates around the torus, so that this is just a gauge transformation; if we take the phases in the square $[-\pi,\pi] \times [-\pi,\pi]$, opposite boundaries are identified under smooth gauge transformations and the twist space is a torus. Under the change of basis (11), Kane and Mele's model Hamiltonian[5] $H = H_0 + H_{SO} + H_R$ (see Section I) becomes (suppressing spinor indices)

$$\begin{aligned} H \to \mathcal{H}(\boldsymbol{\phi}) = &\sum_{\langle ij\rangle} c_i^\dagger \left[t + i\lambda_R(\boldsymbol{s} \times \hat{\boldsymbol{d}}_{ij})_z\right] c_j e^{-i\boldsymbol{\Delta}\cdot\boldsymbol{d}_{ij}} \\ &+ i\lambda_{SO} \sum_{\langle\langle ij\rangle\rangle} \nu_{ij} c_i^\dagger s^z c_j e^{-i\boldsymbol{\Delta}\cdot\boldsymbol{d}_{ij}} \\ &+ \sum_i (\lambda_v \xi_i + w_i) c_i^\dagger c_i, \end{aligned} \quad (12)$$

where again $\boldsymbol{\Delta} = (\phi_x/L_x, \phi_y/L_y)$, $\boldsymbol{d}_{ij}$ is still the vector $i \to j$, and we have added a random term for on-site disorder, w_i, drawn from the Gaussian distribution of zero mean and standard deviation σ_w. It is now clear that under time reversal,

$$\Theta\mathcal{H}(-\boldsymbol{\phi})\Theta^{-1} = \mathcal{H}(\boldsymbol{\phi}), \quad (13)$$

since the extra phase factors in $\mathcal{H}$ change sign under complex conjugation.[†] Since this directly parallels equation (7), the $\mathbb{Z}_2$ invariant D of the Brillouin zone passes directly to twist space:

$$D_\phi = \frac{1}{2\pi}\left[\oint_{\partial(ETZ)} d\boldsymbol{\phi}\cdot\boldsymbol{\mathcal{A}} - \int_{ETZ} d^2\phi\, \mathcal{F}\right] \quad \mathrm{mod}\ 2, \quad (14)$$

with ETZ for Effective Twist Zone, i.e., ETZ $= \{\boldsymbol{\phi}\,|\,0 \le \phi_x \le \pi,\ -\pi < \phi_y \le \pi\}$. Note that there is an independent Chern parity D_ϕ for each Kramers-degenerate band pair separated from the rest of the spectrum by a gap at all $\boldsymbol{\phi}$: for such an isolated pair, $\boldsymbol{\mathcal{A}}$ and $\mathcal{F}$ are defined as in (8) and after, with $\boldsymbol{k} \to \boldsymbol{\phi}$ and $u \to \chi$.

In order to make contact with the band-structure definition, we note that if there is no disorder in the Hamiltonian (i.e., $\sigma_w = 0$), there are discrete translational symmetries within the $L_x \times L_y$ supercell that induce additional non-Kramers degeneracies at some points in twist space. With such degeneracies only the total Chern parity of all the degenerate states is well-defined. For example, in the clean graphene model, there are only two separated sets of states, even though the dimension of the Bloch Hamiltonian increases as L_x and L_y increase.

[†] Note that this means $\mathcal{H}$ is not generically time-reversal invariant. Indeed, there are only four boundary conditions that respect time-reversal, namely those for which ψ picks up a real phase upon translation around each cycle of the honeycomb lattice. These four correspond to the TRIM of the clean system in the calculation of D.

Each of these sets has its own Chern parity, consistent with the counting of $\mathbb{Z}_2$ invariants in the band structure (note that the two invariants are not independent because of a zero sum rule).[11] As in the band-structure case, the phase of a physical system is determined by the sum of Chern parities for occupied band pairs. Disorder, as discussed in the following section, breaks all degeneracies resulting from translational invariance, leaving only separated band pairs, each of which has its own Chern parity. We now discuss to what extent Chern parities can be connected to observable quantities in a finite system.

C. Charge pumping cycles in time-reversal-invariant systems

The above definition was motivated by thinking of a finite system as a (possibly large) unit cell of an infinite periodic lattice. The finite system was defined to be in the topological insulator phase according to whether the infinite lattice is in the topological insulator phase. Since the original identification of this phase was by numerical observation of stable edge states in finite systems, it seems clear that there should be a direct way to detect the topological insulator in the finite system.

The total Chern number in a finite IQHE system can be interpreted as measuring the number of charges pumped when the flux through one noncontractible circle on the torus increases adiabatically by one flux quantum. Briefly, one of the boundary phases corresponds to this driving flux, and the average over the other can be shown to yield the pumped charge.[28] The idea of "$\mathbb{Z}_2$ pumping" suggested by Fu and Kane[10] is the following: in a finite cylinder with boundaries, the operation of increasing the phase ϕ_x (in the periodic direction, around the cylinder) from 0 to π, corresponding to a magnetic flux of one-half flux quantum through the cylinder, has the following effect in the topological insulator. The values $\phi_x = 0$ and $\phi_x = \pi$ are special because, unlike general values, they are consistent with time-reversal invariance. At these special fluxes there are gapless states at the Fermi level that are localized near the edges because of the bulk gap. Fermi statistics requires that these states lie in Kramers doublets. If at zero flux, the Kramers doublet at one edge is partially occupied (has one state occupied), then the operation of changing the flux changes its occupancy to either double or zero occupancy. Since total charge is conserved, this requires a flow of $\mathbb{Z}_2$ from one boundary to the other. Note that in addition to the average over applied flux, there is effectively an average over the spatial extent of the edge, because the edge states are extended.

We now give an alternate definition of the topological insulator in terms of *cyclic* pumping of ordinary charge. This definition is mathematically equivalent to the definition of II B based on treating the finite system as a supercell. In order to describe a closed pumping cycle, we need to add a second stage to the Fu-Kane process of increasing boundary phase ϕ_x from 0 to π: although both these phases are consistent with time-reversal invariance, physical properties, like the occupancy of an edge doublet, are not identical at these different values of ϕ_x (even at the same ϕ_y). The only requirement on the second stage is essentially that it return the system to its original state *without applying a time-reversal-breaking flux*. The second stage gives a closed pumping cycle that returns the system to its original Hamiltonian, which means that the amount of charge pumped in a specified cycle is a well-defined integer quantity.

Although the number of charges pumped is dependent not only on the first stage but on the second stage as well, whether this is an even or odd number is entirely determined by the first stage, as we now show. A closed pumping cycle is shown in Fig. 3. (As before, we discuss the cycle in terms of variable Hamiltonians in a fixed Hilbert space rather than in terms of changing the boundary conditions on Hilbert space.) The original physical system's ETZ is the first stage of the cycle: the Hamiltonians are functions of ϕ_x from 0 to π and ϕ_y from $-\pi$ to π, with time-reversal constraints that act on the boundary circles at $\phi_x = 0$ and $\phi_x = \pi$. If ϕ_x takes one of these values, then the Hamiltonian at ϕ_y is time-reversal conjugate to that at $\phi'_y = -\phi_y$. The second stage can be any continuous change of the Hamiltonians that takes the $\phi_x = \pi$ system back to the $\phi_x = 0$ system and always satisfies the conjugacy condition between ϕ_y and ϕ'_y. This is the key difference between the second stage and the first stage: for intermediate values $0 < \phi_x < \pi$, there is no such conjugacy condition. The physical interpretation is that the second stage should be possible without introducing flux through the first noncontractible circle.

Now the torus shown in Fig. 3 has one Chern integer for each isolated band pair. Summing over occupied bands gives the amount of charge pumped in the cycle. Although this integer charge depends on the second stage, its parity is solely determined by the first stage, i.e., the physical system. In particular, for the ordinary insulator there is some closed cycle that pumps zero charge, while for the topological insulator there is some closed cycle that pumps unit charge. These results follow from the same proof as for the band structure case in Ref. 11: one shows that the differences in resulting Chern integers between any two second stages are even. The pumping definition gives some physical intuition for the "contractions" introduced there; instead of contracting the EBZ to a sphere, here the ETZ is contracted to a torus by adding the second stage. The technical reason that these two constructions are equivalent is that, since the appropriate spaces of Hamiltonians are contractible (i.e., have $\pi_1 = 0$), the two closed manifolds, the torus and the sphere, both have the same topological invariants, namely integer-valued Chern numbers.

The topological insulator in disordered systems has been studied previously by locating the transition between topological and ordinary insulators as a point or region where the localization length of single-electron

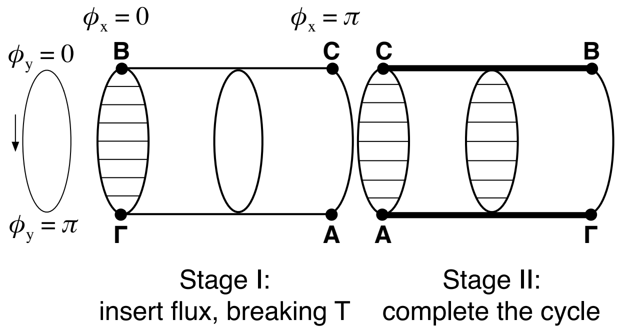

FIG. 3: Graphical representation of charge pumping cycle for Chern parities. The first stage takes place on the ETZ (as in equation (14)), and the flux ϕ_x increases adiabatically from 0 to π. In the second stage the Hamiltonian at $(\phi_x = \pi, \phi_y)$ is adiabatically transported through the space of Hamiltonians to return to the Hamiltonian at $(\phi_x = 0, \phi_y)$. The difference between the second stage and the first is that at every step of the second stage, the Hamiltonians obey the time-reversal conditions required at $\phi_x = 0$ or $\phi_x = \pi$. The bold lines indicate paths along which all Hamiltonians are time-reversal invariant, and the disk with horizontal lines indicates, as before, how pairs of points in the second stage are related by time-reversal.

eigenstates diverges. The existence of the topological insulator is inferred from the existence of this transition region (or alternately from the edge states in the topological insulator phase). In principle this approach is different from ours, in that our definition probes the existence not just of extended states but specifically of extended states that contribute to the pumping of charge, or alternately can be expected to give rise to edge states. The same distinction arises in the quantum Hall effect, where looking for extended states of nonzero Chern number is a more direct probe of quantum Hall physics than considering the inverse participation ratio, for example, which would detect extended states of zero Chern number in addition to topological states. However, as in the quantum Hall case, we find that the phase boundaries from our Chern-parity definition are consistent with those obtained from calculations of the localization length.[14,16]

III. GRAPHENE MODEL AND NUMERICS

A. The phase diagram of the disordered graphene model

It is useful to review some general expectations before applying the definitions of the previous section to study a disordered version of the graphene model. In two-dimensional systems with time-reversal invariance and no spin-orbit coupling, even very weak disorder will localize electron wavefunctions, so that these systems do not ever conduct in the thermodynamic limit. In the presence of a magnetic field, as in the IQHE, there are isolated energies with extended states, but no finite-width range of energies with extended states, and hence no true metallic phase. Hikami *et al.* showed[29] that disordered two-dimensional systems *with spin-orbit coupling* can nevertheless support a metallic phase, referred to as a "symplectic metal". In our case, any metallic phase would presumably appear in a region around the parameter set that closes the (clean) gap, as depicted schematically in Fig. 4.

We also know that at $\lambda_R = 0$ the z component of spin is a good quantum number (s^z commutes with H), so the system reduces to two copies of the Haldane model,[8] which has a quantum Hall plateau transition with no metallic phase.

B. Lattice Implementation

For numerical work we use the algorithm of Fukui and Hatsugai, which we review here.[13] The formula (9) for D requires a gauge choice for the Hamiltonian eigenstates at each $\boldsymbol{\phi}$ on the two boundaries of the half-torus $0 \leq \phi_x \leq \pi$, $-\pi < \phi_y \leq \pi$. That is, the "field strength" $\mathcal{F}$ is gauge invariant, but the gauge potential $\boldsymbol{\mathcal{A}}$ is not. The eigenstates form Kramers pairs related by time reversal, and the gauge choice must respect this constraint.

Now, at the time-reversal invariant points $\boldsymbol{\phi} = (0,0)$, $(0,\pi)$, $(\pi,0)$, and (π,π), the solid points in Fig. 5, the spectrum is degenerate, with two states at each energy. The gauge condition requires that Θ interchange the two with a phase factor $e^{\pm i\pi/2}$ (so that $\Theta^2 = -1$ as required for single-fermion states). Numerical diagonalization will not, in general, return eigenvectors that obey this condition, but we can force them to do so as follows: choosing one of the two members of each Kramers pair at energy $\varepsilon_{2n-1} = \varepsilon_{2n}$ and calling that vector χ_{2n-1}, we discard the other and replace it by

$$\chi_{2n} = \Theta\chi_{2n-1}. \tag{15}$$

On the rest of the boundary, eigenvectors can be chosen freely on $0 < \phi_y < \pi$. On $-\pi < \phi_y < 0$ the algorithm takes

$$\chi_n(-\boldsymbol{\phi}) = \Theta\chi_n(\boldsymbol{\phi}). \tag{16}$$

In summary, the algorithm leaves alone the results of numerical diagonalization at all the gray points in Fig. 5, and by hand enforces the gauge condition on the rest of ∂(ETZ).‡

‡ The points (ϕ_x, π) and $(\phi_x, -\pi)$ are not physically distinct; as pointed out earlier, a gauge transformation relates $\mathcal{H}(\phi_x, -\pi)$ to $\mathcal{H}(\phi_x, \pi)$ as given, for example. We therefore impose $\mathcal{H}(\phi_x, -\pi) \equiv \mathcal{H}(\phi_x, \pi)$ in the calculation. Alternatively, we could introduce boundary terms along $\phi_y = \pi, -\pi$ to compensate for the discrepancy.

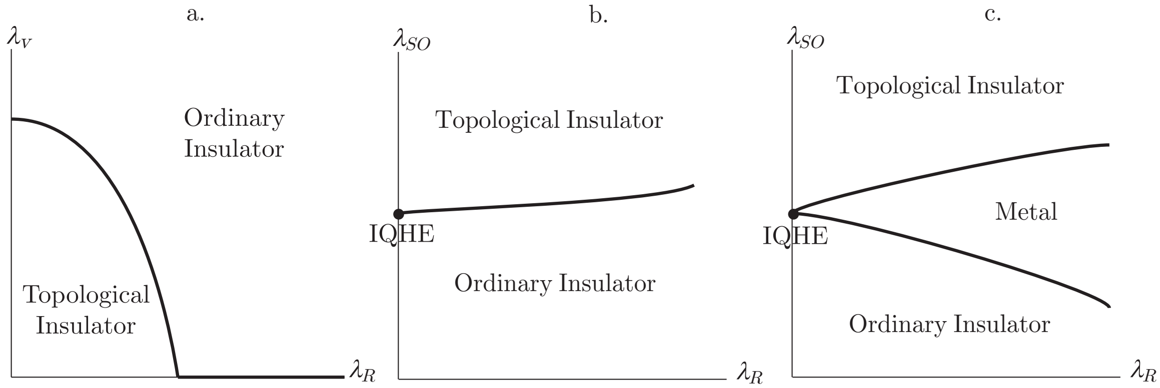

FIG. 4: Schematic phase diagram. (a) Phase diagram for a clean system, with fixed t and $\lambda_{SO} \neq 0$ (after Kane and Mele[6]). (b) Again a clean system, now with fixed t and $\lambda_v \neq 0$. (c) The expected form of the phase diagram at nonzero disorder (we run all simulations at fixed λ_v). The phase boundary in (b) opens up into a metallic phase, closing only when $\lambda_R = 0$, where there should be an IQHE transition.

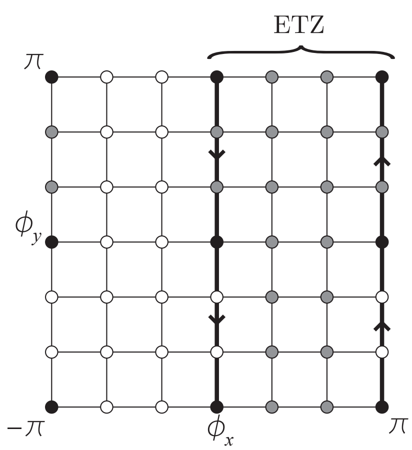

FIG. 5: Twist Space. The bold lines indicate the boundaries of the "Effective Twist Zone", the region we integrate (or sum) over to calculate the Chern parity. The arrows indicate the direction to perform the sum over the boundary terms, and the lattice sites in gray indicate those for which the Hamiltonian eigenvectors are independently specified. That is, time reversal symmetry determines the eigenvectors on the white sites once those at the gray sites are found.

With the eigenstates fixed at each point on the ETZ, we follow Fukui and Hatsugai and construct $U(1)$ parallel transporters on the links, as

$$U_x(\boldsymbol{\phi}) = \frac{g_x}{|g_x|},\ g_x = \det \boldsymbol{\chi}^\dagger(\boldsymbol{\phi})\boldsymbol{\chi}(\boldsymbol{\phi}+\hat{\boldsymbol{x}}) \qquad (17)$$

and U_y similarly. Like $\boldsymbol{u}$ in (8), $\boldsymbol{\chi}$ is a matrix built from occupied state vectors, and $\hat{\boldsymbol{x}}$ translates by one link in the ϕ_x direction. In the continuum limit g should approach a pure phase, but for non-zero lattice spacing it will in general have $|g| \leq 1$, since the occupied subspace of interest will not embed in the total Hilbert space in the same way at every lattice point.

In the end, the only retained information will be the variation of the relative phases (hence the definition of U), which can be captured by choosing a lattice constant so small that the phase field varies slowly over one link. However, the scale of variation will presumably differ for different disorder realizations, and we would like a way to diagnose this and throw out those realizations for which fast variation makes the calculation unreliable. The phase is periodic, and if it were to wind through 2π over the distance of one link the algorithm would miss this fact, so the relative phase will not provide a good diagnostic. As a proxy, we choose to cull out disorder realizations that result in small determinants in most simulations, since a small overlap between adjacent occupied eigenspaces indicates rapid variation. Of course, this filtering could introduce a selection bias into the results; these effects are within the statistical uncertainty of our analysis (in particular, within the error bars of Fig. 9).

Associated with the transporter on each link is a gauge potential $A_{x,y} = \log U_{x,y}$. This A is pure imaginary, and the logarithm is defined to return the branch $A/i \in (-\pi, \pi)$. Associated with the transport around each plaquette is a flux

$$F(\boldsymbol{\phi}) = \log U_x(\boldsymbol{\phi})U_y(\boldsymbol{\phi}+\hat{\boldsymbol{x}})U_x^{-1}(\boldsymbol{\phi}+\hat{\boldsymbol{y}})U_y^{-1}(\boldsymbol{\phi}), \quad (18)$$

again satisfying $F/i \in (-\pi, \pi)$. With these definitions in hand, the lattice $\mathbb{Z}_2$ invariant corresponding to D_ϕ (14) is

$$D_L = \frac{1}{2\pi i}\left[\sum_{|\in\partial(ETZ)} A_y - \sum_{\square\in ETZ} F\right] \mod 2. \quad (19)$$

Of course, the sum over the boundary should have the same orientation as the corresponding contour integral in (14); in Fig. 5, this means the sum on the left boundary should carry a minus sign, following the arrows. As mentioned previously, this formalism has the desirable

property of guaranteeing that $D_L = 0$ or 1, but using too coarse a mesh can return the wrong value.

C. Numerical Results

As noted after equation (14), the number of nondegenerate Kramers pairs in a disordered system will generically be extensive, and we can use equation (19) to calculate the Chern parity of each pair separately. Fig. 6 shows the results of such calculations on a 4×6 lattice to present a picture of the three phases: normal insulator, symplectic metal, and topological insulator. In all phases, there are Kramers pairs with $D_\phi = 1$ in the lower half of the energy spectrum, as indicated by the bars. However, in the normal insulator (λ_{SO} well below the transition, or gap closure) there are an *even* number of such pairs in the occupied half of all realizations, so that the overall Chern parity is even. The presence of a very small number of realizations (2 of 215 for a 6×8 lattice) with odd parity indicates either the tail of the disorder-broadened transition or, more likely, that the weak filter applied for these simulations failed to catch all realizations with rapidly varying link variables.

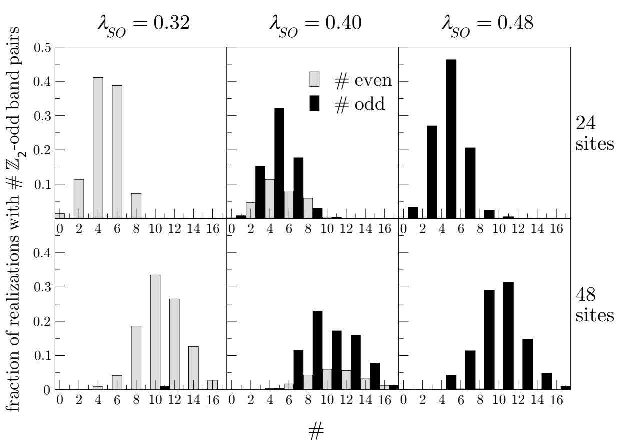

FIG. 6: Distribution of Chern parities for band pairs. The bar heights represent the fraction of disorder realizations (out of ~ 200 trials) that have a given number # of band pairs with $D_L = 1$ in the occupied (half-filled) subspace. Those with an even number # will have an overall $D_L = 0$, those with an odd number will have overall $D_L = 1$. All these simulations were done with $t = -1$, $\lambda_R = \lambda_v = 1$, and $\sigma_w = 0.3$. Reading across, λ_{SO} increases and the system transitions from all realizations having even parity at small λ_{SO} to odd parity at large λ_{SO}. Reading down, doubling the system size doubles the total number of Kramers pairs and roughly doubles the number of $\mathbb{Z}_2$-odd pairs.

When λ_{SO} is large, almost all realizations have an odd number of $\mathbb{Z}_2$-odd Kramers pairs (as with the small-λ_{SO} case, 2 realizations do not follow the rule), and in the region near the transition of the clean system there are instances of both types. The presence of disorder causes the "gap" to close at different values of λ_{SO} for different realizations, and also at different energies. The latter fact means that extended states are present throughout a finite spread of energies, while the former means that the metallic state exists over a finite region of parameter space.

For an integer quantum Hall system, Yang and Bhatt[19] have shown how to extract the localization length exponent ν from such calculations, in their case the Chern numbers of Landau sublevels. Specifically, sublevels with non-zero Chern number contain extended states, which should only occur at isolated energies in the IQHE. Therefore, the number N_c of such sublevels should decrease as the system size N_s increases, and in fact $\langle N_c\rangle \propto N_s^{1-1/2\nu}$, where $\langle\rangle$ indicates an average over disorder realizations. A similar approach for the QSH system here should reveal $\langle N_D\rangle \propto N_s$, where N_D is the number of $\mathbb{Z}_2$-odd Kramers pairs, since we expect a stable metallic band of energies in the thermodynamic limit (as observed by Obuse *et al.*[14] and Onoda *et al.*[16]). With enough data and large enough systems, finite-size-induced broadening of the edges of this band should also make ν accessible via a subleading term in the scaling. We find that larger system sizes require a finer mesh in twist space for these Kramers-pair-resolved simulations to return stable results, so that the requirements quickly outstrip our resources. Nevertheless, comparison of the two rows in Fig. 6 indicates that the location of the mean roughly doubles. This is what would be expected for the middle case given the above considerations; the mean for the topological insulator (the right-hand panels in Fig. 6) should grow more slowly, as in the case studied by Yang and Bhatt.

The total phase of the system, given by the total Chern parity, is more relevant to possible measurements than the Chern parity of each Kramers pair. The former maintains its meaning if we consider the ground state wave function of the many-electron system, formed as a Slater determinant of the single-electron states we use here. There is also a computational benefit to calculating D_ϕ for the whole occupied subspace rather than for individual pairs, as well — at larger system sizes, and also at stronger disorder, the link variables for the half-filled subspace vary much more slowly than those for the individual Kramers pairs, making the calculation more robust. For these reasons the remaining plots in this paper depict only the Chern parity of the half-filled system.

To show that a metallic region of non-zero extent in parameter space exists in the thermodynamic limit, we need to verify that the mixed-phase region does not shrink to zero as we increase the system size. Figure 7(a) shows that as we make the system larger, the transition region certainly does not get narrower, and in fact the largest system size seems to have the broadest transition.

We can quantify the scaling of the metallic region's width with system size by assuming a simple one-parameter scaling form for the curves in Fig. 7(a) and

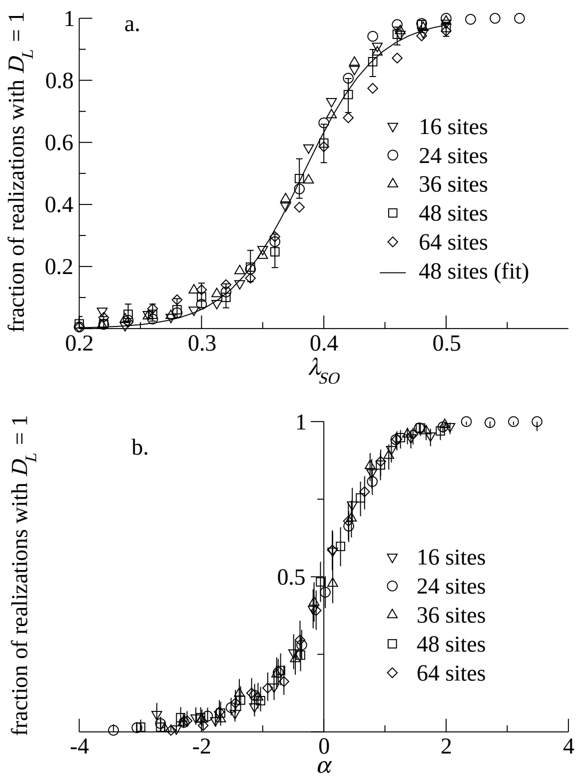

FIG. 7: (a) At finite λ_R, the "metallic" region persists as the system size grows, and even broadens in the case shown here ($t = -1$, $\lambda_R = \lambda_v = \sigma_w = 1$). We identify the metallic region as those values of λ_{SO} for which some, but not all, disorder realizations have $D_L = 1$. As explained in the text, the fit to the simulation data (shown only for 6×8 systems here) has the form $(\tanh \alpha + 1)/2$, with $\alpha = m(\lambda_{SO} - \lambda^*_{SO})$. The error bars represent 95% confidence intervals assuming a binomial distribution of outcomes for each λ_{SO}. (b) The scaling collapse of the data in (a), based on the best fit m and λ^*_{SO}) for each system size.

defining the width of the curve to be proportional to the reciprocal of the maximum slope: width $\sim$ 1/slope. With sufficient data, one could expand the approximately linear region near the middle of the transition in a power series with a few coefficients as fit parameters. Since our simulation data are limited, we opt instead to assume the form $(\tanh \alpha(s) + 1)/2$, which has roughly the right shape. If $\alpha = m(\lambda_{SO} - \lambda^*_{SO})$, then m is exactly the maximum slope we want. Figure 7(b) plots the data versus the best-fit scaling variable α for each system size; the points appear to fill out a smooth curve, justifying the scaling hypothesis. The best fit for m and λ^*_{SO} is determined by minimizing a weighted χ^2 statistic.[30] In particular, we assume that the results ($D_L = \pm 1$) of independent simulations at a fixed parameter set are distributed binomially, and that for each system size there are the same number of trials at each value of λ_{SO} (which is roughly true). In that case, the variance of the distribution can be estimated as $\sigma^2 \propto p(1-p)$,[31] where p is the fraction of disorder realizations returning $D_L = 1$, and then $1/\sigma^2$ is an appropriate weight for the statistical test. The error bars in Figs. 7 and 8 are also assigned based on a binomial model of the data.[30]

By contrast, at $\lambda_R = 0$ the Hamiltonian (5) reduces to two copies of the Haldane model and so should exhibit the quantum Hall plateau transition, which looks like a step function at zero temperature in the thermodynamic limit. In Fig. 8(a) the width of the transition region shrinks as the system size grows, consistent with the prediction.

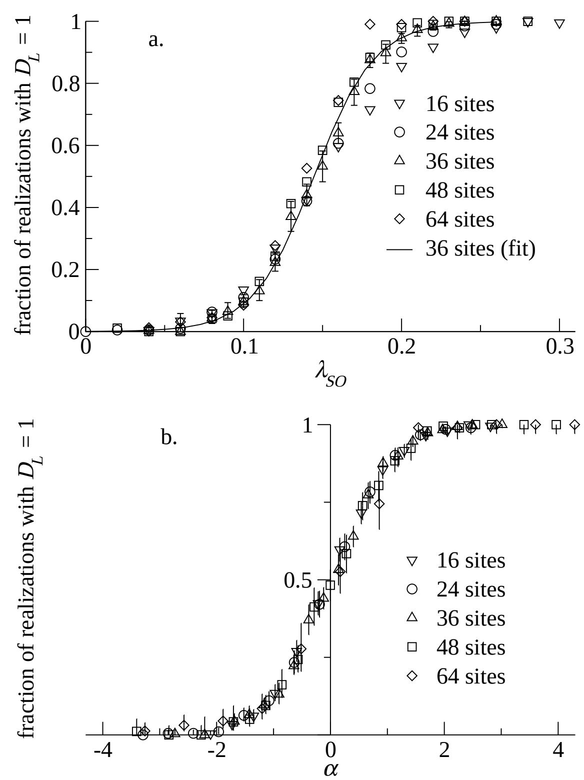

FIG. 8: (a) At $\lambda_R = 0$, the metallic region gets narrower as the system size increases. The fit is to a 6×6 system and $\lambda_R = 0$; as in Fig. 7, $t = -1$, $\lambda_v = \sigma_w = 1$. (b) The scaling variable α again comes from fitting to a tanh, and error bars represent 95% confidence intervals.

More quantitatively, Pruisken[32] has shown in a renormalization group framework that the functional form of the crossover for the IQHE looks like

$$p(L, B) = f(\alpha), \quad \alpha \propto L^{1/\nu}(B - B^*), \tag{20}$$

where L is the linear size of the finite sample, B is the applied magnetic field, and ν is again the localization length exponent. The function p could be either the longitudinal or transverse conductivity in the IQHE. Therefore, $p(\infty, B^* \pm \varepsilon) = f(\pm\infty)$, i.e., the transition is sharp in

the thermodynamic limit. In our system the analogous parameter to B is λ_{SO}: in the Haldane model, the spin-orbit coupling breaks time-reversal invariance locally, like B does in the IQHE. The width of the transition region in B is governed by the way the Landau level energies respond to changing B, and the width of the transition region in our model (for fixed λ_v) is determined by the response of the gap to changing λ_{SO}. Since both responses are linear, we expect that the appropriate scaling variable will be $\alpha \propto L^{1/\nu}(\lambda_{SO} - \lambda_{SO}^*)$.

This form would allow us to extract the exponent ν from the scaling of the maximum slope with system size for large systems (there are corrections at small system sizes). Again making the fit to a tanh described above, Fig. 8(b) shows the scaling form, and Fig. 9 shows the scaling of width with linear system size. In particular, a regression gives $1/\nu = 0.78 \pm 0.03$, to be compared with the accepted value of $1/\nu \approx 0.42$. That is far from good agreement, but the observed scaling should not be taken as implying a new universality class. (For reference, the network-model work by Obuse *et al.*[14] found $1/\nu \approx 0.37$, and Onoda *et al.*[16] recently found a value $1/\nu \approx 0.63$) First, there should be finite-size corrections to the simple scaling assumed for the small systems considered here. Second, the scaling form is in principle different for different geometries, and the simulations were done for systems of varying aspect ratio (1 to 1.5). Nevertheless, it is clear that the qualitative behavior at $\lambda_R = 0$ is as expected, showing no metallic phase, and the behavior at $\lambda_R = 1$ is consistent with the presence of a metallic region.

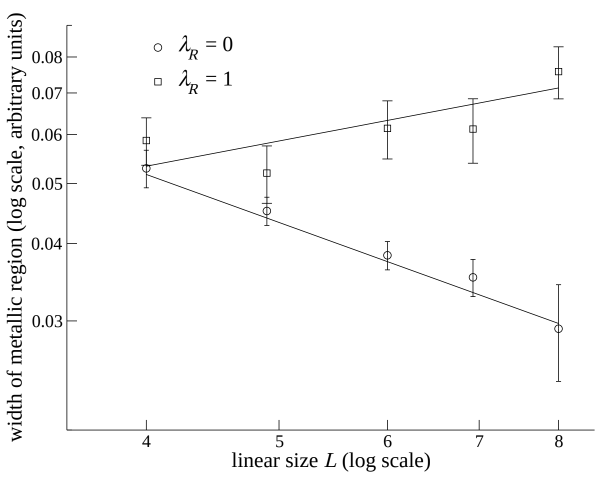

FIG. 9: Width of the metallic transition region as a function of linear size L. The width is determined by the reciprocal of the maximum slope of the simulation data in Figures 7 and 8, and the linear size is taken to be the square root of the number of sites.

Finally, by varying λ_R and noting the λ_{SO} values that mark the edges of the transition region for each λ_R, we can map out the phase diagram of the Hamiltonian

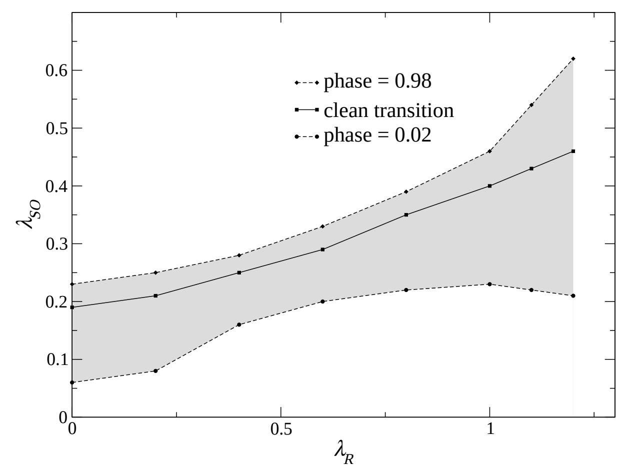

FIG. 10: Approximate width of metallic region for a 4×6 lattice (fixed $\lambda_v = \sigma_w = 1$). The dashed curves indicate the parameter values at which 98% and 2% of the disorder realizations have $D_L = 1$. This underestimates the true width of the "metallic" region but hopefully avoids some amount of the inevitable error due to small system size. This diagram should offer a reasonable approximation to the thermodynamic (infinite-system) phase diagram away from $\lambda_R = 0$. At $\lambda_R = 0$, there should be no metallic phase, but a sharp transition between the two insulating phases in the thermodynamic limit. Given the plots in Fig. 7, it appears that finite size effects also reduce the width of the metallic phase at large λ_R.

(5), as in Fig. 10. The widths obtained this way are in rough agreement with the scaling analysis outlined above, which returns the behavior of the width and not its normalization. As the results in Figures 7 and 8 show, this phase diagram will overestimate the width of the metallic phase at $\lambda_R = 0$, which is really zero. Together, these simulations confirm the expectation of Fig. 4 within the accuracy of our computational methods.

IV. SUMMARY

Previous work[5,6,10,11,33] defined a $\mathbb{Z}_2$ topological invariant in infinite lattices that is similar to the TKNN invariant for the integer quantum Hall effect.[20] In disordered systems with boundaries, Fu and Kane[10] defined a topological invariant in terms of pumping of the occupancy of Kramers-degenerate edge states. We have given a definition of a topological invariant valid for disordered systems without boundary, i.e., without appeal to edge states. The "Chern parity" can be thought of as describing either a finite system with boundary phases or an arbitrarily large supercell in an infinite lattice system with well defined wavevector. A physical effect of Chern parity is that it determines whether the amount of charge pumped in a certain type of closed pumping cycle is even or odd. The "ordinary" and "topological"

insulator phases can be distinguished by this invariant as long as many-body effects do not prevent description of the ground state as a Slater determinant. Chern parity in the spin quantum Hall effect is the natural generalization of Chern number in the integer quantum Hall effect.

In a disordered system, the only degeneracies expected to survive are Kramers degeneracies at time-reversal invariant values of the boundary phases. In this case, each pair of states related by Kramers degeneracies can be assigned its own Chern parity, and the overall Chern parity of all occupied state pairs determines the observable phase. The lattice algorithm for $\mathbb{Z}_2$ topological invariants laid out by Fukui and Hatsugai[13] allows numerical identification of the topological insulator phase in disordered QSH systems. Implementing this algorithm for the specific graphene model Hamiltonian of Kane and Mele[6] with added on-site disorder, we observe the ordinary and topological insulator phases in simulations. While the number of "odd" pairs (state pairs with odd Chern parity) varies with the disorder realization, there is an even number of odd pairs in the ordinary insulator, and an odd number of odd pairs in the topological insulator.

We find that a metallic phase opens up between the two insulating phases for generic spin-orbit coupling. This agrees with the prediction of Hikami *et al.*[29] that spin-orbit coupling can protect a 2D metallic phase from disorder, and confirms the simulation results of Onoda *et al.*[16] and Obuse *et al.*[14] The methods in this paper could in principle be used to study the three-dimensional case and confirm the argument in Ref. 34 that only one of the four invariants of a band structure[11,21,23] is stable to disorder. While there is now strong evidence[14] that the phase transitions in the 2D QSHE are, except for special points, in the previously studied symplectic metal-insulator class, there is as yet no numerical study of the phase transitions in three-dimensional topological insulators.

Acknowledgments

The authors thank L. Balents, H. Haggard, D.-H. Lee, C. L. Kane, and A. Vishwanath for useful comments and acknowledge support from NSF DMR-0238760 and an NDSEG fellowship (A. E.). Numerical work took place on a cluster donated by the IBM SUR program.

[1] S. Murakami, N. Nagaosa, and S.-C. Zhang, Science **301**, 1348 (2003).
[2] J. Sinova et al., Phys. Rev. Lett. **92**, 126603 (2004).
[3] Y. K. Kato, R. C. Myers, A. C. Gossard, and D. D. Awschalom, Science **306**, 1910 (2004).
[4] J. Wunderlich, B. Kaestner, J. Sinova, and T. Jungwirth, Phys. Rev. Lett. **94**, 047204 (2005).
[5] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).
[6] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).
[7] M. I. D'yakonov and V. I. Perel, JETP Lett. **13**, 467 (1971).
[8] F. D. M. Haldane, Phys. Rev. Lett. **61**, 2015 (1988).
[9] D. N. Sheng, Z. Y. Weng, L. Sheng, and F. D. M. Haldane, Phys. Rev. Lett. **97**, 036808 (2006).
[10] L. Fu and C. L. Kane, Phys. Rev. B **74**, 195312 (2006).
[11] J. E. Moore and L. Balents, Phys. Rev. B **75**, 121306 (2007).
[12] T. Fukui and Y. Hatsugai, Phys. Rev. B **75**, 121403 (2007).
[13] T. Fukui and Y. Hatsugai, J. Phys. Soc. Jpn. **76**, 053702 (2007).
[14] H. Obuse, A. Furusaki, S. Ryu, and C. Mudry, cond-mat/0702199.
[15] J. T. Chalker and P. D. Coddington, J. Phys. C: Solid State Phys. **21**, 2665 (1988).
[16] M. Onoda, Y. Avishai, and N. Nagaosa, Phys. Rev. Lett **98**, 076802 (2007).
[17] R. B. Laughlin, Phys. Rev. Lett. **52**, 2304 (1984).
[18] D. E. Khmelnitskii, Phys. Lett. A **106**, 182 (1984).
[19] K. Yang and R. N. Bhatt, Phys. Rev. Lett. **76**, 1316 (1996).
[20] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, Phys. Rev. Lett. **49**, 405 (1982).
[21] R. Roy, cond-mat/0607531.
[22] B. A. Bernevig and H.-D. Chen, cond-mat/0611766.
[23] L. Fu, C. L. Kane, and E. J. Mele, Phys. Rev. Lett **98**, 106803 (2007).
[24] J. E. Avron, R. Seiler, and B. Simon, Phys. Rev. Lett. **51**, 51 (1983).
[25] J. J. Sakurai, *Modern Quantum Mechanics* (Addison-Wesley, 1994), chap. 4.4.
[26] Q. Niu, D. J. Thouless, and Y.-S. Wu, Phys. Rev. B **31**, 3372 (1985).
[27] J. E. Avron and R. Seiler, Phys. Rev. Lett. **54**, 259 (1985).
[28] Q. Niu and D. J. Thouless, Phys. Rev. B **35**, 2188 (1987).
[29] S. Hikami, A. I. Larkin, and Y. Nagaoka, Prog. Theor. Phys. **63**, 707 (1980).
[30] W.-M. Yao et al., Journal of Physics G **33** (2006), §32.
[31] G. D'Agostini, Rept. Prog. Phys. **63**, 1383 (2003).
[32] A. M. M. Pruisken, Phys. Rev. Lett. **61**, 1297 (1988).
[33] R. Roy, cond-mat/0604211.
[34] L. Fu and C. L. Kane, cond-mat/0611341.